SYNOPSIS

Investigation of Remote Sensing Images Using Deep Neural Network Analysis

Submitted by

(Nadeem Akhtar Riyaz Ahmad)

(SAP ID – 70821016002)

Supervisor(s)

Dr. Manish Mandloi

MUKESH PATEL SCHOOL OF TECHNOLOGY MANAGEMENT & ENGINEERING
SVKM's NARSEE MONJEE INSTITUTE OF MANAGEMENT STUDIES
(Declared as Deemed-to-be University Under Section 3 of the UGC Act, 1956)
V. L. Mehta Road, Vile Parle (West)
MUMBAI –400056

(2022-23)

ABSTRACT

Semantic segmentation of remote sensing images is a challenging task that involves the automatic labeling of each pixel in an image with its corresponding class. The goal is to accurately classify the land cover types within the image, which is essential for applications, like environmental monitoring, urban planning, and disaster management. This task is complicated by the variability of remote sensing data, which can come from a variety of sources, including satellite imagery, and aerial photography. The deep learning architecture has shown promising results which led to the development of a range of novel architectures and training techniques. However, there are still many challenges that need to be addressed, such as dealing with highly imbalanced classes, handling large-scale datasets, dealing with occluded regions and differentiating region of interest with same texture region, and training with minimum computation cost. This study addresses the segmentation of natural, non-natural resources using deep learning architectures. The natural resources like forest and non-natural resources like roads and buildings are extracted from Massachusetts, Shanghai, Birmingham, Boston, Inria and Vaihingen dataset. Two model are proposed for occluded road detection. The first model use zooming and shrinking paths in the model architecture to extract the road features followed by geometrical shape analysis for filtering the non-road objects. The second architecture use a dense residual Segnet for extraction of occluded road. The two proposed architectures achieve the state-of-art performance on occluded road and achieves better performance metrics as compare to other methods. A boundary assisted building detection is also proposed in the study. The proposed method fuses DenseResSegnet with building boundary detection module and channel attention module. The model is tested on Inria and Massachusetts building dataset. For multiclass semantic segmentation, a deep dense residual bottleneck layer is proposed which uses asymmetric convolution layers. This third layer of this bottleneck uses dilated convolution to capture the objects shape. The proposed layers are parallel processed at different dilation in deep layers of the encoder decoder based architecture. The asymmetric convolution layer with parallel process reduces the computation cost, hence memory. It boost the batch size and smooths learning curve and time. The model achieves state-of-art performance on Deep globe forest dataset and Vaihingen dataset.

***Keywords*:** Remote sensing, Deep neural networks, convolutional neural network, encoder decoder architecture, pyramid pooling, dilated inception, road detection, building detection, semantic segmentation, post processing, channel attention, edge detection module, asymmetrical convolution layer, dilated convolution, attention module, dense residual connections, pattern recognition.

TABLE OF CONTENTS

ABSTRACT 2

CHAPTER 1 ..... ***5***

INTRODUCTION **5**

1.1. RESEARCH PROBLEM 5
1.2. RESEARCH GAP 5
1.3. RESEARCH OBJECTIVES 5

CHAPTER 2 ..... ***6***

BACKGROUND **6**

2.1. SEMANTIC SEGMENTATION OF REMOTE SENSING IMAGES 6

CHAPTER 3 ..... ***7***

LITERTURE REVIEW **7**

3.1. Literature review on Semantic Segmentation. 7

CHAPTER 4 ..... ***9***

CONTRIBUTIONS **9**

4.1. CONTRIBUTION 1- Semantic Segmentation of narrow and wide road under occlusion. 9
4.1.1 Simulation Environment and Datasets. 10
4.1.2 Results and Discussion. 12
4.1.3 Ablation Study. 17
4.2. CONTRIBUTION 2- Semantic Segmentation Road with light weight architecture. 19
4.2.1 DenseResSegnet. 19
4.2.2 Ablation Study of DenseResSegnet 21
4.3. CONTRIBUTION 3- Semantic Segmentation building on Inria and Massachusetts dataset. 22
4.3.1 Data set for simulation 23
4.3.1 Results and Discussion on Building Detection. 24
4.4. CONTRIBUTION 4- Semantic Segmentation multiclass on Vaihingen dataset. 25
4.4.1 Vaihingen and Deepglobe Dataset. 27
4.4.2 Results and Discussion on Forest Detection. 30

CHAPTER 5 ..... ***32***

SUMMARY AND CONCLUSIONS **32**

CHAPTER 6 ..... ***34***

FUTURE WORK **34**

PUBLICATION **35**

LIST OF FIGURES

Figure 1: Proposed basic model (BM) architecture for road segmentation 9
Figure 2: Dilated Inception Module 10
Figure 3: Pyramid Pooling Module 10
Figure 4: Bounding rectangle area for (a) Straight-road; (b) curved-road [25] 11
Figure 5: (a) Segmented image and (b) Post processed image 12
Figure 6: Simulation qualitative analysis on Massachusetts road dataset 13
Figure 7: Simulation qualitative analysis on Birmingham dataset 15
Figure 8: Simulation qualitative analysis on Boston dataset. 16
Figure 9: Simulation qualitative analysis on Shanghai dataset. 17
Figure 10: Ablation of dice metric with datasets 18
Figure 11: Ablation of quality metric with datasets 18
Figure 12: Segnet architecture [27] 19
Figure 13: proposed bottleneck layer [42] 19
Figure 14: Simulation qualitative analysis on Massachusetts road dataset 21
Figure 15: DenseResSegnet with building boundary detection and channel attention module 22
Figure 16: Squeeze and Excitation module as channel attention module [44]. 23
Figure 17: Simulation results of proposed network Inria dataset. 24
Figure 18: Simulation results on Massachusetts building dataset. Purple and green color indicates the FP and FN. 25
Figure 19: bottleneck layers with weight size 26
Figure 20: Down sampling scheme 27
Figure 21: Proposed bottleneck layer 27
Figure 22: layers with dilated convolutions 27
Figure 23: Proposed architecture for semantic segmentation 28
Figure 24: Intersection over Union [26] 29
Figure 25: Visual comparison results of proposed network with ERF net 29
Figure 26: Simulation results on DeepGlobe forest data set. 31

LIST OF TABLES

Table 1: Literature review on semantic segmentation 7
Table 2. Analysis of proposed network with other methods on Massachusetts dataset. 13
Table 3: Comparison with published statistics on Massachusetts road dataset 14
Table 4: Analysis of proposed network with other methods on Birmingham dataset. 14
Table 5: Analysis of proposed network with other methods on Boston dataset. 15
Table 6: Analysis of proposed network with other methods on Shanghai dataset. 16
Table 7: Analysis of proposed network with Boston, Birmingham and Shanghai dataset 17
Table 8. AS of the proposed method on Massachusetts dataset. 18
Table 9: AS of the proposed method on Boston dataset. 18
Table 10: AS of the proposed method on Birmingham dataset. 18
Table 11: AS of the proposed method on Birmingham dataset. 19
Table 12: Comparative analysis of DR Segnet with state of art methods. 20
Table 13: Comparison with published statistics on Massachusetts road dataset 21
Table 14: Ablation study of proposed method with constant and variable filters and dilated convolution 22
Table 15: Performance metrics on Inria dataset 24
Table 16: Class wise performance metrics on Inria dataset 24
Table 17: Performance metrics on Massachusetts dataset 25
Table 18: Class wise performance metrics on Massachusetts dataset 25
Table 19: Number of trainable parameters required for bottleneck layer 26
Table 20: Performance metrics on Massachusetts dataset 30
Table 21: Class wise performance metrics on Massachusetts dataset 30
Table 22: Performance metrics on Deep Globe dataset 30
Table 23: Class wise performance metrics on Deep Globe dataset 30

ABBREVIATION NOTATION AND NOMENCLATURE

RS	*Remote Sensing*
DL	*Deep Learning*
CNN	*Convolutional Neural Network*
FCN	*Fully Convolutional Network*
F	*Number of Convolution Filter*
S	*Stride*
BM	*Basic Module*
PPM	*Pyramid Pooling Module*
DI	*Dilated Inception*
S_{MER}	*Area of Minimum Bounding Rectangle*
TP	*True Positive*
TN	*True Negative*
FP	*False Positive*
FN	*False Negative*
A	*Area*
C	*Complex Rate*
F	*Fullness Ratio*
IoU	*Intersection over Union*
ResU-net	*Residual U-net*
AS	*Ablation Study*
DR	*Dense Residual*
DRB	*Dense Residual Block*
BN	*Batch Normalization*
ReLU	*Rectified Linear Unit*
GPU	*Graphical Processing Unit*
C	*Concatenation*
U	*Up-sampling*
+	*Aggregation*
SE	*Squeeze and Excitation*
CRF	*Conditional Random Fields*

CHAPTER 1
INTRODUCTION

1.1. RESEARCH PROBLEM

The use of aerial images for various applications such as urban planning, disaster response, and agriculture has increased significantly in recent years. Semantic segmentation, the task of assigning each pixel in an image to a specific object or class, is a crucial component in many of these applications. However, semantic segmentation of aerial images poses several challenges such as complex urban scenes with varying object sizes and shapes, highly variable object textures and colors, and significant class imbalances. Thus, research problem is how to develop robust and accurate models that can handle these challenges and achieve state-of-the-art performance. Several studies have been conducted; however, these methods have limitations in capturing complex spatial structures and textures, handling class imbalance. The research problem addressed in this study is how to develop effective and efficient semantic segmentation models for aerial images that capture complex spatial structures and textures, with multiple classes, such as buildings, vegetation, water bodies, and roads, particularly in scenarios where the classes are highly similar or have significant overlaps.

1.2. RESEARCH GAP

1. Most of the semantic segmentation network is improved with increasing the depth of the model, design of interconnection layer and modified loss function.

2. Increasing the depth of the network by introducing the interlayer connection such as attention module increases the memory requirement of the computing device and hence need elongated training time.

3. The training time and number of features in encoder of model architecture boost the memory consumption and hence the training time.

4. With increase in number of classes in dataset degrades the model performance (IoU).

5. Binary class algorithms performance degrades for occluded road in shadow region due to tree and building shadow, further output is affected by the nearby parking lots.

1.3. RESEARCH OBJECTIVES

1. Detection of natural information like agricultural area, water resources and desert landscape from the multispectral and aerial images using convolutional neural network.

2. Detection of non-natural information like road, building and dams from multispectral and aerial images using convolutional neural network

3. Detection of natural and non-natural information from multispectral and aerial images using convolutional neural network

CHAPTER 2
BACKGROUND

2.1. SEMANTIC SEGMENTATION OF REMOTE SENSING IMAGES

Semantic segmentation is a crucial task in computer vision that involves assigning labels to each pixel in an image based on its similar properties, such as color, texture, or object class. The primary objective of this task is to label each pixel in an image with its corresponding object class. Advances in semantic segmentation techniques have enabled their application to remote sensing problems, which are complex and data-rich. Examples of remote sensing applications include environmental monitoring, crop cover and type analysis, tree species identification in forests, building classification, and land use analysis in urban spaces. Multispectral remote sensing images can provide detailed information about the Earth's surface that is invisible to the human eye. These images consist of multiple bands, each corresponding to a different region of the electromagnetic spectrum. They are valuable in identifying and labeling natural and man-made objects such as buildings, roads, trees, water bodies, and landmarks. This information can be useful in various applications such as urban planning, disaster management, and environmental monitoring. The semantic segmentation of multispectral remote sensing images presents unique challenges due to the high spatial and spectral variability of the images. Accurately identifying the boundaries of different land cover types, such as water bodies, forests, and urban areas, while accounting for variations in illumination, atmospheric conditions, and sensor noise is a challenging task. This type of segmentation is particularly useful in analyzing large areas, such as entire cities or regions, and monitoring changes in land cover over time.

To address these challenges, a range of techniques, from traditional image processing methods to state-of-the-art deep learning algorithms, have been developed. One of the primary challenges in semantic segmentation of multispectral remote sensing images is accurately distinguishing between similar land cover types. Deep learning techniques, such as convolutional neural networks (CNNs), have been increasingly applied to this task. These approaches involve training a model on a large dataset of labeled images, allowing the algorithm to learn features that can be used to accurately classify pixels. These techniques have been successful in various applications, including precision agriculture, land cover mapping, and disaster management, outperforming traditional image processing methods.

CHAPTER 3
LITERTURE REVIEW

3.1. Literature review on Semantic Segmentation.

Table 1: Literature review on semantic segmentation

Ref	Methodology	Features	Remarks
[1]	CNN	round objects are extracted accurately.	Obtaining pixel intensities in shadow regions incurs higher costs.
[2]	ResNet101-v2 with pyramid pooling	The framework provided has shown superior performance and has proven to be effective.	Due to the substantial spectral similarities between impermeable surfaces and buildings, the segmentation accuracy is restricted.
[3]	Multitask CNN	The approach has presented a systematic and adaptable framework for achieving efficient segmentation outcomes.	The method does not retain the complex geometrical characteristics required for accurate segmentation.
[4]	Cascaded framework	The approach has enhanced the predictions by incorporating object boundaries and eliminates the isolated false positives.	computational cost is high
[5]	Dual-path densely connected convolutional networks	The method circumvents the vanishing gradient issue and improves of inter-layer information using dense connections.	Despite its limitations, the method boasts faster training and testing times.
[6]	CAM-DFCN	The model has improved segmentation results by using effective feature selection.	The proposed CAM-DFCN did not exhibit a significant improvement in performance when compared to the CNN+RF+CRF approach.
[7]	FCN	This model has the capability to recover lost data, resulting in improved robustness and accuracy.	This method is not suitable for shadowed areas.
[8]	ESPCN and UNet	The approach has improved segmentation and robustness of the model.	Needs more training samples
[9]	AD-LinkNet	It has improved efficiency of segmentation.	The model performance degrades on various road interruptions.
[10]	FCN	The approach improves segmentation efficiency compared to conventional methods while incurring less overhead.	Integrating multiple tasks degrades the segmentation efficiency.
[11]	CNN	This model focuses on representative features and has enriched the semantic information.	The model's performance is limited due to difficulties in processing high-level features.
[12]	GCN	This model effectively captures complex features and improves segmentation while also addressing the issue of data scarcity.	To improve accuracy, the approach may benefit from incorporating optimization techniques and semantic labelling strategies.
[13]	Modified InceptionV-4 network (DAPN)	This method has robust generalization capabilities.	While consistent performance was achieved on the Potsdam dataset, but model performance degrades on the Vaihingen dataset.
[14]	Nested dilated residual block (NDRB)	The approach provides accurate object boundaries and labeling for complex scenes.	The evaluation of per-class accuracy is not performed, which may limit the assessment of the approach's effectiveness.
[15]	HMANet	The approach captures global contextual information to improve segmentation efficiency, while also enhancing the self-attention	The method requires a significant amount of memory consumption.

		mechanism and reducing feature redundancy.	
[16]	SegNet	The approach exhibits superior building extraction capabilities for medium- and high-sized buildings.	Fails in identifying the small building.
[17]	ALRNet	The approach selects the most non-redundant and representative features, hence improves the model performance.	This approach have high computational inefficiency.
[18]	Integrated algorithm (encoder-decoder CNN structures SegNet with index pooling and UNet)	This method integrates features of CNN and U-net for improving the results of semantic segmentation.	This approach has degraded performance for some classes in comparison with other algorithm.
[19]	UNet	This approach is capable of precise segmentation of tree canopies, and it can also solve complex problems in environments such as agricultural production.	Model fail in addressing the problem of densely merged and closely located false positives.
[20]	Aerial-BiSeNet	The model has achieves a good stability between speed and accuracy, demonstrating better efficiency and performance on both datasets	The model's representation ability is weak, and it has a high level of complexity.
[21]	GAMNet	The integration module has been improved to enhance efficiency, resulting in accurate results with accurate boundaries, for small objects.	The model faces a misclassification problem in shaded areas, leading to confusion.
[22]	GCN	The model effectively restores ground object boundaries and reduces pixel-level noise.	However, this model does not take advantage of spatial correlation information to better interpret remote-sensing images.

CHAPTER 4
CONTRIBUTIONS

4.1. CONTRIBUTION 1- Semantic Segmentation of narrow and wide road under occlusion.

The proposed architecture for segmentation utilizes 15 convolution layers with 3×3 filters, followed by a batch normalization and ReLU layer. The BM architecture consists of up-sampling (yellow block), down-sampling (orange block), and bridge sections (pink block) in Figure 1. By utilizing both up-sampling and down-sampling paths, the BM architecture can analyze images at multiple resolutions. The up-sampling path increases image size and halves feature size, while the down-sampling path reduces image size by a factor of 2 and doubles the feature depth. This process extracts road features for segmentation quickly. The BM architecture uses transposed and strided convolutions for up-sampling and down-sampling, respectively. Thick and thin features are extracted at different resolution scales, and residual skip connections are used to connect the up and down sampled features.

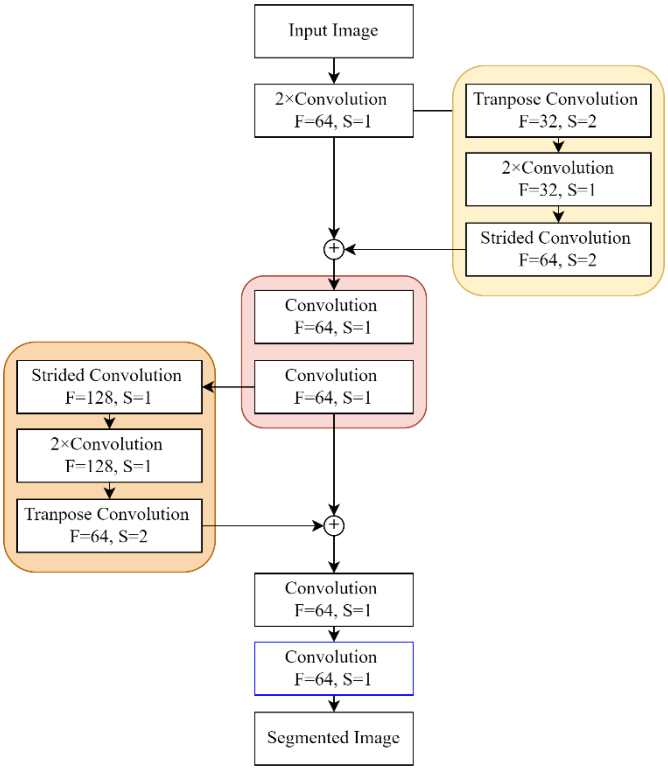

Figure 1: Proposed basic model (BM) architecture for road segmentation

The prediction accuracy in image processing is heavily influenced by the field of view, which refers to the amount of contextual information available for analysis. Insufficient field of view can result in suboptimal outcomes, especially when dealing with partial information about structures such as buildings or roads. This lack of information makes it difficult to distinguish between objects with similar textural features. In addition, predictions for pixels located at the

edges of patches, also known as marginal pixels, can be unreliable due to the limited field of view. However, dilated convolutions can be used to extract multi-scale contextual information while maintaining the same computational cost and increasing the field of view. This helps in generating large-scale feature maps with rich spatial information for object recognition. The DI module shown in Figure 2 uses dilated convolutions in the up-sampling and down-sampling paths of the BM architecture, with 3 parallel convolution paths and different dilation rates of 4, 8, and 16.

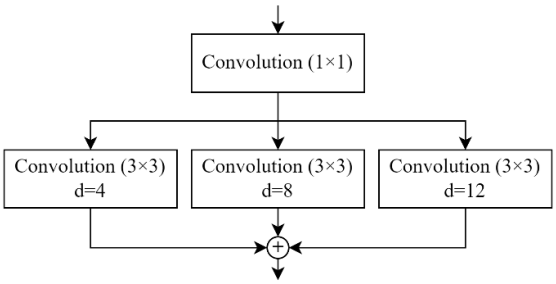

Figure 2: Dilated Inception Module

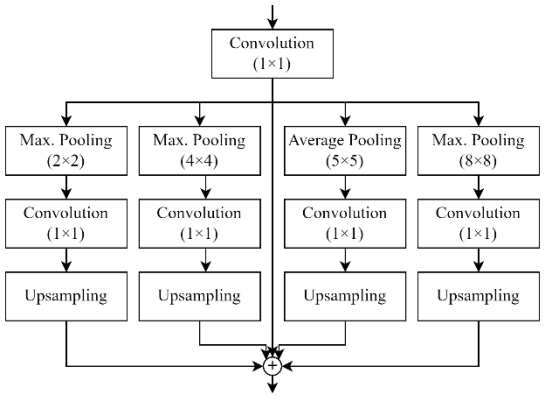

Figure 3: Pyramid Pooling Module.

The BM architecture's bridge component, depicted in the pink block, is replaced with the PP module. The PP module employs multiscale pooling to extract deep features and captures road information by applying average sub region pooling with window sizes of 2×2, 4×4, 5×5, and 8×8, as shown in Figure 3. The resulting features are convolved and up sampled using bilinear interpolation, and then concatenated with the previous layer's features.

4.1.1 Simulation Environment and Datasets.

Four datasets are used for the experiment: Massachusetts [23], Boston, Birmingham, and Shanghai [24]. The Boston, Birmingham, and Shanghai datasets were collected from Google Earth and are publicly available. The Massachusetts dataset has 1171 aerial images of the Massachusetts state with road and non-road labeled classes. The Boston image has a resolution of 0.4411 m/pixel and a size of 23104×23552 pixels. The Birmingham image has a resolution of 0.3640 m/pixel and a size of 22272×22464 pixels. The Shanghai image has a resolution of 0.5109 m/pixel and a size of 16768×16640 pixels. Shanghai has significantly different ground object distribution with many narrow roads between relatively dense and tall buildings compared to Boston and Birmingham. We have randomly chosen 335 patches of 512×512 size from Massachusetts dataset and 1900 patches of 256×256 from Shanghai, Birmingham and Boston image for training the proposed model. Models are simulated using a core i7, 32GB

RAM, and RTX3070 GPU with dedicated 8 GB internal memory. The models are trained for 50 epochs using progressive learning rate with a drop factor of 22. The initial learning rate is 0.001 with Adam optimizer and batch size of 6.

Evaluation Metrics.

The metrics used for road detection are correctness, completeness and quality are given by (1), (2), (3) and (4) respectively.

$$\text{Correctness} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100 \tag{1}$$

$$\text{Completness} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \tag{2}$$

$$\text{Quality} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \times 100 \tag{3}$$

$$\text{Dice} = \frac{2 \times \text{Correctness} \times \text{Completness}}{\text{Correctness} + \text{Completness}} \times 100 \tag{4}$$

TP indicates that the road pixels are correctly identified; FN indicates that the road pixels are falsely identified while false positive indicates that the non-road pixels are identified as road pixels. The geometrical parameters used for filtering are; area (A), Complex Rate (C) and Fullness Ratio (F) [25]. The shape complexity of the object in binary image is defined by (5) and (6) using perimeter (P) and A.

$$\text{Complex Rate } = \frac{\text{P}^2}{\text{A}} \tag{5}$$

$$\text{Fullness Ratio } = \frac{\text{A}}{\text{S}_{\text{MER}}} \tag{6}$$

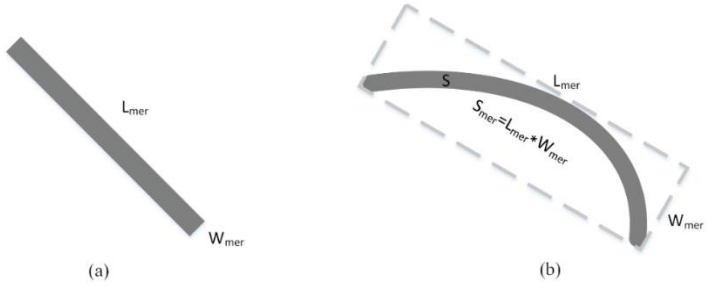

Figure 4: Bounding rectangle area for (a) Straight-road; (b) curved-road [25]

$$(\text{A} > A_{\text{T}})\&\&(\text{C} > \text{C}_{\text{T}})\&\&(\text{F} < \text{F}_{\text{T}}) \tag{7}$$

The non-road not satisfying the (7) are filtered as non-road objects. The results of post processing are shown in Figure 5.

(a) Segmented Image

(b) Post processed image

Figure 5: (a) Segmented image and (b) Post processed image

Loss function.

There are several loss functions that can be used for semantic segmentation, and the choice of loss function depends on the specific requirements of the task. Some common loss functions used in semantic segmentation include cross entropy, dice, focal and intersection over union (IoU) loss. The systematic effect of loss function is analyzed by [43] on remote sensing data. For simulation purpose we have used focal loss function. The choice of loss function in semantic segmentation depends on task requirements, including class balance and number of classes. Road detection data has a class imbalance with 10% road pixels, affecting model performance. Loss function selection is crucial for convergence and optimization of model performance. Focal loss handles class imbalance by learning hard samples and down weighting easy samples.

4.1.2 Results and Discussion.

Massachusetts Dataset: The comparative analysis of the proposed method and the other methods are shown in Figure 6, Figure 7, Figure 8 and Figure 9. The visual results of Massachusetts data set are shown in Figure 6. This shows that Segnet [27], U-net [28], ResU-net [29] and FCN [30] are falsely identifying the non-road as road object in Figure 6 (e). The proposed network identifies the road non-road as noise. The Figure 6 (d) contains the occluded road by shadows of tree. FCN detects it partially as compare to proposed network. The point where more than two roads are meeting is road junction, this point may have different appearance than the roads due to road circle etc. The proposed method is able to identify the road junctions as well while other methods fail in detection the roads junction points. The visual result comparison is shown in Figure 6. The Segnet fails in identifying one road segments as shown in Segnet Figure 6 (a, d). The U-net is unable to identify the occluded roads are shown in U-net Figure 6 (b, c, d). The FCN fails in identifying the occluded road in Figure 6 (a). The track road with rough surface is identified by the proposed method while Segnet, U-net, ResU-net and FCN fails in identifying this as a non-road object. Table 2 shows the quantitative comparative analysis of proposed method and other methods for comparison. These numeric

values represent the mean of 32 testing images. Segnet achieves highest completeness among the methods considered. Segnet preserves continuity of road due to the pooling indices are transferred to the decoder part. The quality and dice of proposed method is highest among the considered methods.

Table 2. Analysis of proposed network with other methods on Massachusetts dataset.

Network	Completeness	Correctness	Quality	Dice
Segnet [27]	**87.15**	*84.65*	73.47	*84.65*
U-net [28]	75.33	82.93	65.09	78.65
ResU-net [29]	82.77	81.31	69.44	81.90
FCN [30]	*85.46*	84.11	*73.60*	73.60
Proposed	85.41	**88.00**	**76.52**	**86.66**

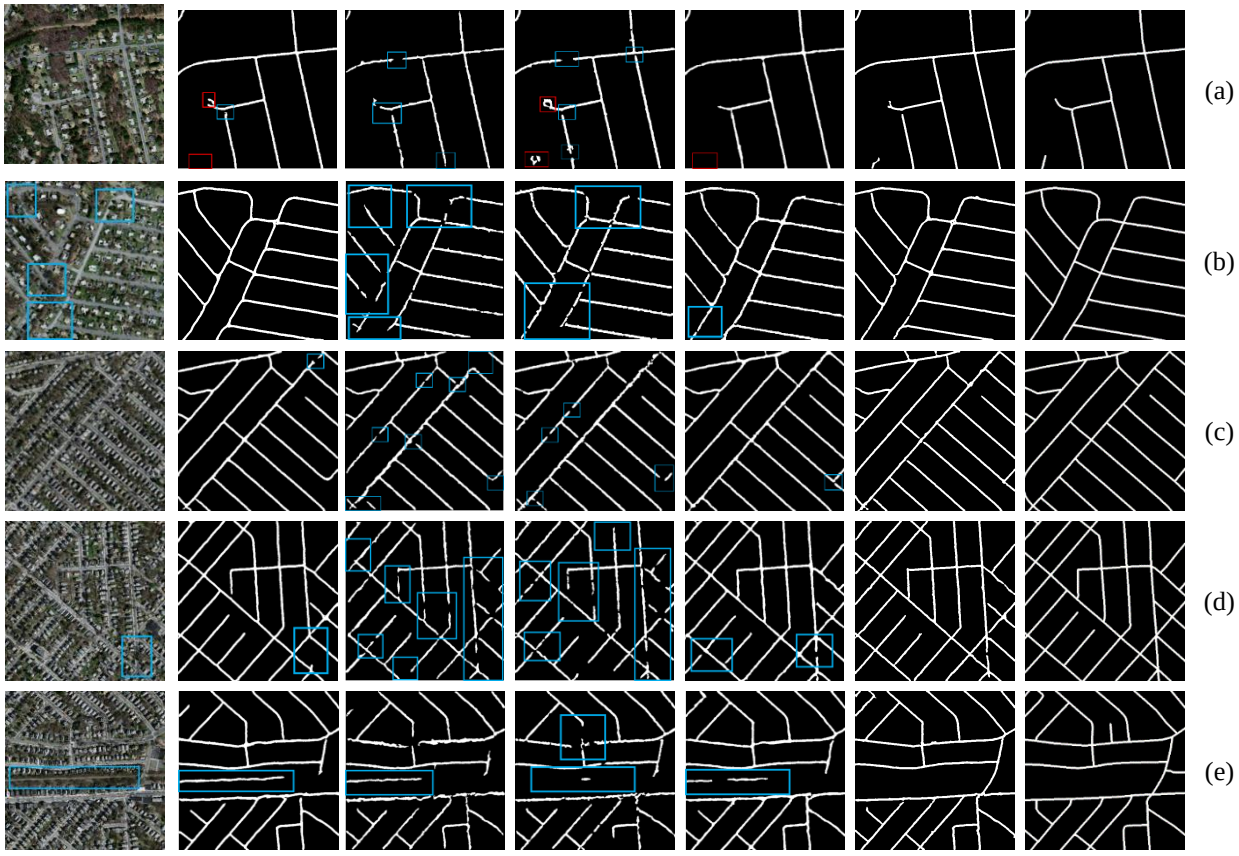

Figure 6: Simulation qualitative analysis on Massachusetts road dataset

Table 3: Comparison with published statistics on Massachusetts road dataset

Ref.	Methodology	Completeness	Dice	Correctness	Quality
[31]	Cascaded Attention Dense U	76.55	77.89	79.45	64.12
[32]	ResNet with Multi Task	72.54	78.13	84.65	64.11
[33]	Inter-Layer Self-Attention	56.66	66.19	79.58	-
[34]	Encoder-Decoder and DCRF	82.2	*84.6*	*87.1*	-
[35]	Dense U Net	70.41	74.07	78.25	*74.47*
[36]	Link Net	76.10	75.17	74.30	60.32
[37]	EU Net with Ensemble learning and Post Processing	81.31	80.45	80.71	68.56
[38]	Densely Connected Spatial	71.97	76.59	-	62.48
[39]	DA-Road Net	77.25	78.19	79.16	64.19
[40]	Deep FR TransNet	*83.72*	-	83.72	62.86
[41]	Richer U Net	-	-	-	62.77
Proposed Method		**85.41**	**86.66**	**88.00**	**76.62**

Table 3 shows the comparison with other methods published statistics. In this proposed methods outperforms the others methods

Birmingham Dataset The visual results of the Birmingham dataset are presented in Figure 7, where (a) and (b) exhibit tree shadows and (d) displays a road junction with a different spatial signature. While Segnet fails to distinguish these features, the proposed network is able to accurately identify them. Notably, the proposed network correctly identifies the tree shadow region as road in (a) and (b), which U-net fails to do. Table 4 provides a quantitative comparative analysis of the Birmingham dataset based on the mean of 190 testing images. Segnet achieves the highest completeness and dice scores, while U-net has the highest correctness among the evaluated methods. The proposed methods are the second-highest for quality and dice scores in the road detection task, among the considered methods.

Table 4: Analysis of proposed network with other methods on Birmingham dataset.

Network	Completeness	Correctness	Quality	Dice
Segnet [27]	**86.72**	77.02	*68.38*	**80.63**
U-net [28]	57.02	**87.71**	51.78	66.36
ResU-net [29]	*83.98*	70.53	61.14	75.23
FCN [30]	83.87	*82.01*	**70.30**	81.99
Proposed	82.82	79.54	67.72	*80.57*

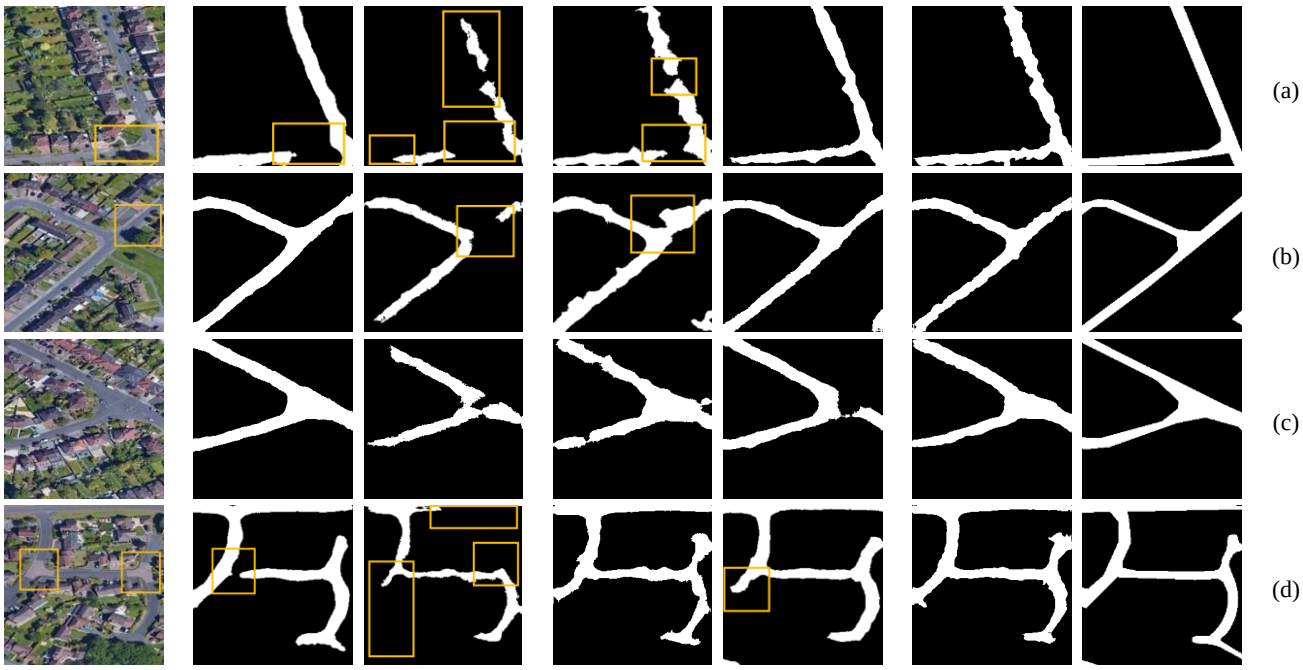

Figure 7: Simulation qualitative analysis on Birmingham dataset

Boston Dataset: The visual and quantitative comparative analysis of the Boston dataset using the proposed and other methods are presented in Figure 8 and Table 5, respectively. In this dataset, the roads are surrounded by building top roofs that have a similar spectral signature to the road. Figure 8 (a), (c), and (e) depict similar textures of road with neighboring building top roofs, and the roads are shadowed by trees. However, the proposed method is able to effectively discriminate between the road and neighboring building top roofs and accurately identify the shadowed road as well. In Table 5, the numeric values represent the mean of 190 testing images. Segnet achieves the highest completeness score, while the proposed network achieves the highest correctness, quality, and dice scores among the evaluated methods for the Boston dataset.

Table 5: Analysis of proposed network with other methods on Boston dataset.

Network	Completeness	Correctness	Quality	Dice
Segnet [27]	**86.30**	88.27	77.07	86.83
U-net [28]	78.52	88.37	70.82	82.53
ResU-net [29]	65.80	75.63	53.24	68.76
FCN [30]	*85.49*	*89.27*	*77.26*	*86.99*
Proposed	85.46	**90.02**	**77.82**	**87.44**

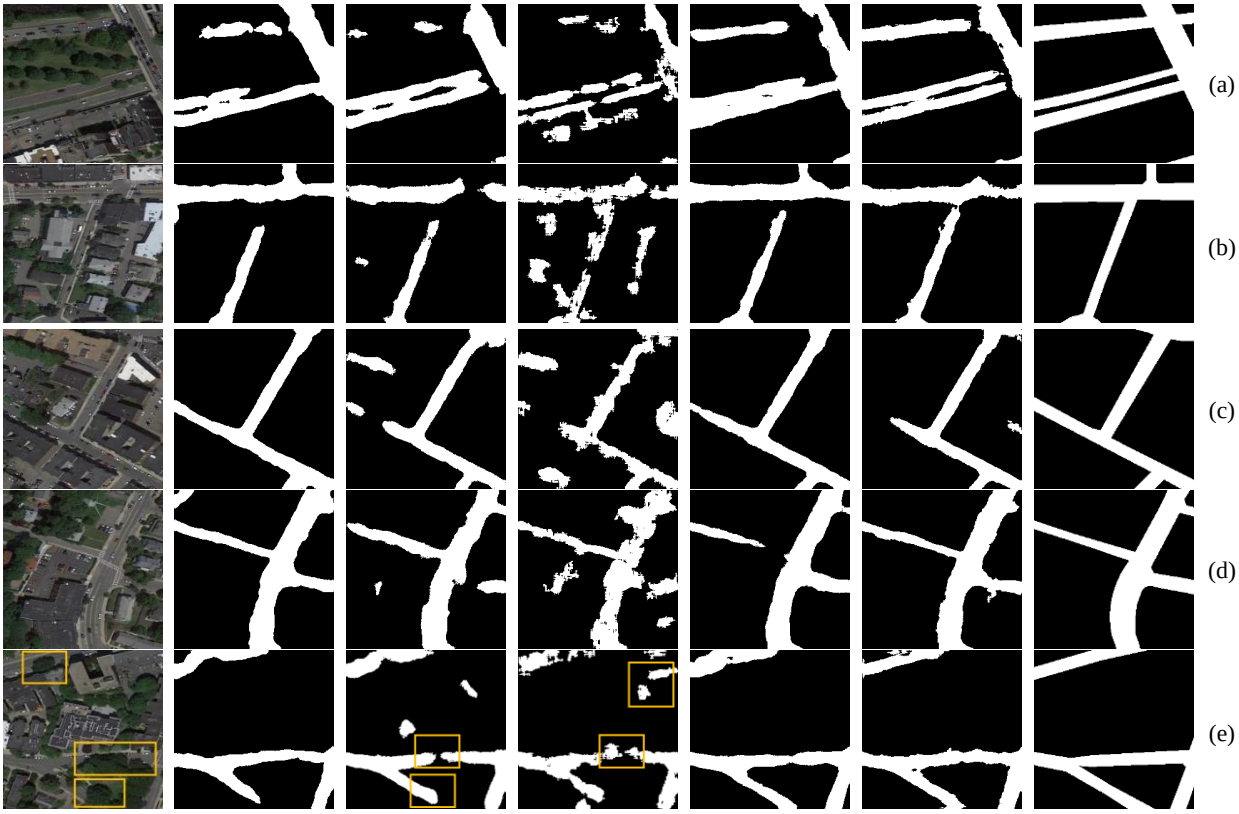

Figure 8: Simulation qualitative analysis on Boston dataset.

Shanghai Dataset: Figure 9 and Table 6 shows the visual and quantitative comparative analysis on proposed and other method considered for analysis on Shanghai data set. Shanghai data set has relatively dense and tall buildings, with many narrow roads between them as compare to the Boston, Birmingham and Massachusetts dataset. The roads in the testing images appear very narrow surrounded by trees, further the road appearance have same texture of surroundings such as building top roofs and river (see Figure 9 (d) and (c)). The visual results shows that the proposed network extracts roads efficiently as compared to the other method considered. The road appears narrower as compare to testing image. In this, Segnet falsely identifies the building top roof as road (see Figure 9 (d) and (c)), further Segnet fails in identifying the road width surrounded by the tree. Segnet has worst performance on Figure 9 (a). The quantitative analysis in Table 6 shows that proposed network has highest completeness, correctness, quality and dice among the considered methods. These numeric values in Table 6 represent the mean of 190 testing images.

Table 6: Analysis of proposed network with other methods on Shanghai dataset.

Network	Completeness	Correctness	Quality	Dice
Segnet [27]	66.163	70.85	51.89	65.21
U-net [28]	67.210	82.28	58.37	72.39
ResU-net [29]	64.864	79.62	55.14	69.91

FCN [30]	*77.344*	*85.24*	*67.79*	*80.18*
Proposed	**79.450**	**86.55**	**70.09**	**82.05**

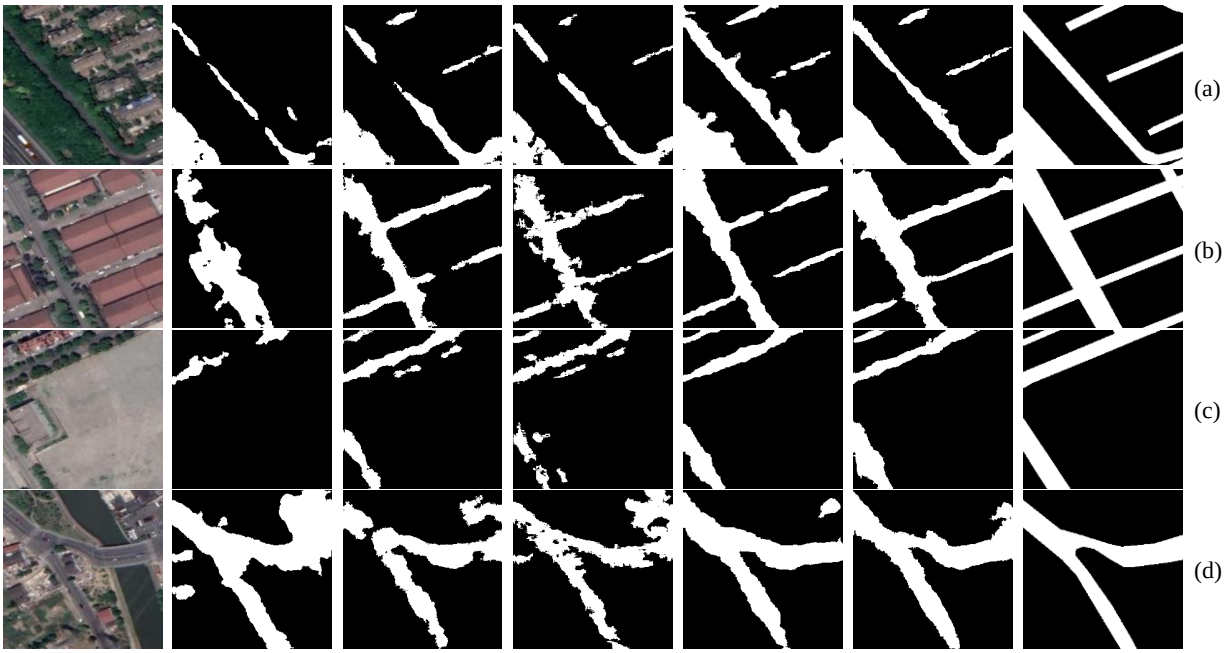

Figure 9: Simulation qualitative analysis on Shanghai dataset.

The proposed method is compared with GAMS net Table 7. In this our network outperforms with GAMS net in terms of dice and quality metric of road extraction.

Table 7: Analysis of proposed network with Boston, Birmingham and Shanghai dataset

Network / Dataset	Boston		Birmingham		Shanghai	
	Dice	Quality	Dice	Quality	Dice	Quality
GAMS-Net [24]	77.37	63.09	72.38	56.71	62.30	45.25
Proposed	**87.44**	**77.83**	**80.57**	**67.72**	**82.06**	**70.09**

4.1.3 Ablation Study.

To verify the effectiveness of the proposed method we have simulated the ablation study (AS) with basic model architecture. The proposed method is composed of DI and PP module. Table 8 Table 9, Table 10 and Table 11 shows the ablation of proposed methods on the dataset considered.

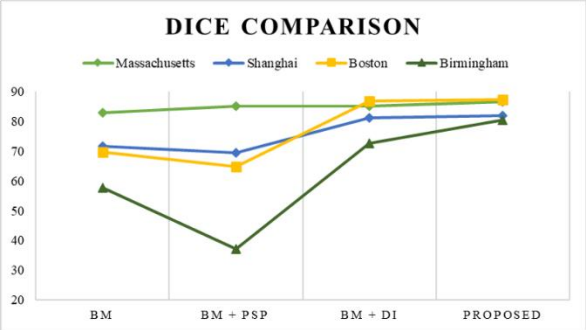

Figure 10: Ablation of dice metric with datasets

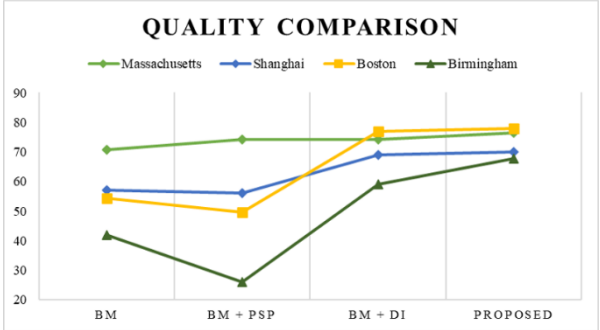

Figure 11: Ablation of quality metric with datasets

To evaluate the effectiveness of the proposed method, we conducted an ablation study (AS) using a basic model architecture. The proposed method consists of the DI and PP modules. Tables 6, 7, 8, and 9 show the ablation results of the proposed method on the considered datasets. The completeness of the proposed model was enhanced by 7, 14, 21, and 34 percent, respectively, compared to the basic model for Massachusetts, Shanghai, Boston, and Birmingham. The Dice coefficient and quality of the proposed model were also improved, with enhancements of 2, 1, 1, and 8 percent for the Dice coefficient, and 2, 1, 1, and 9 percent for the quality, respectively, as shown in Figure 10 and Figure 11. The results indicate a clear improvement in quality for the Birmingham and Boston datasets, where the basic model had poor capability of detecting continuity in road segments.

Table 8. AS of the proposed method on Massachusetts dataset.

Network	Completeness	Correctness	Quality	Dice
BM	78.21	*88.25*	70.79	82.85
BM + PSP	81.42	**89.49**	*74.32*	*85.21*
BM + DI	82.62	87.96	74.17	85.11
Proposed	**85.41**	88.00	**76.52**	**86.66**

Table 9: AS of the proposed method on Boston dataset.

Network	Completeness	Correctness	Quality	Dice
BM	64.14	79.54	54.38	71.68
BM + PSP	55.43	84.54	49.65	69.59
BM + DI	85.43	88.92	*76.97*	*81.13*
Proposed	**85.46**	**90.02**	**77.82**	**82.05**

Table 10: AS of the proposed method on Birmingham dataset.

Network	Completeness	Correctness	Quality	Dice
BM	48.70	77.54	42.03	57.83
BM + PSP	31.25	67.81	25.99	37.25
BM + DI	67.82	84.54	*59.16*	*72.73*
Proposed	**82.82**	**79.54**	**67.72**	**80.57**

Table 11: AS of the proposed method on Shanghai dataset.

Network	Completeness	Correctness	Quality	Dice
BM	65.41	82.82	57.02	71.68
BM + PSP	63.72	85.81	55.98	69.59
BM + DI	77.90	**86.59**	*69.04*	*81.13*
Proposed	**79.45**	*86.55*	**70.09**	**82.05**

4.2. CONTRIBUTION 2- Semantic Segmentation Road with light weight architecture.

4.2.1 DenseResSegnet.

In this we have proposed a bottleneck layer (See Figure 13) termed as Dense Residual Block (DRB) which uses multilevel; convolution, residual and dense connection.

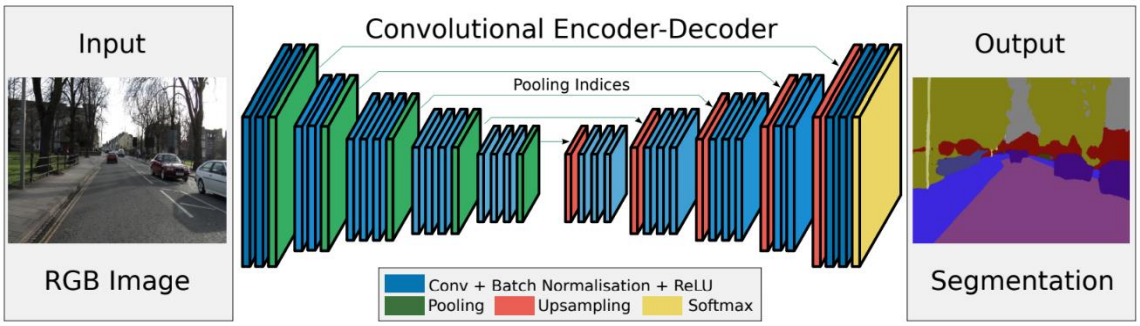

Figure 12: Segnet architecture [27]

The Segnet's convolution layer shown in Figure 12 has been replaced with a DRB as illustrated in Figure 13. The DRB incorporates 1×1 and 3×3 convolution layers along with residual connections, resulting in the aggregation of more contextual information. The last layer features undergo a 1×1 convolution followed by two 3×3 convolutions with residual connection. The dense connection extracts more features from different convolution levels, while the residual connection enhances the network's prediction capability. To ensure optimal performance, each convolution layer in DRB is followed by BN and ReLU. The use of DRB significantly enhances the segmentation ability of Segnet and improves the continuity of road maps in situations where occlusion is caused by tree and building shadows.

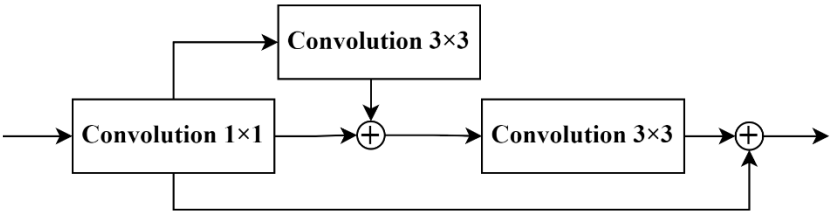

Figure 13: proposed bottleneck layer [42]

The proposed DR Segnet with 64 filters in each stage is trained and tested with state of art

methods like Segnet [27], U-net [28], ResU-net [29] and FCN [30] used for road detection along with post processing.

Hyper parameter: For simulation we have used focal entropy loss with progressive learning rate with 100 epoch, batch size of 14, learning rate drop factor 0.1 with learning rate drop at 30 epochs, early stopping at with 7 epochs and initial learning rate of 0.001.

Table 12: Comparative analysis of DR Segnet with state of art methods.

Network	Completeness	Correctness	Quality	Dice
Segnet [27]	76.15	**91.79**	71.28	83.19
U-net [28]	**87.42**	77.88	70.01	82.24
ResU-net [29]	82.61	86.28	72.96	84.33
FCN [30]	*85.86*	83.17	*73.13*	*84.44*
Proposed	82.94	*87.87*	**74.41**	**85.30**

The inclusion of DR block in the basic Segnet has led to a significant improvement in performance (see Table 12). Our proposed approach was compared with other road detection methods, and it was observed that the use of DRB to aggregate contextual information further enhances road detection performance. Among the compared methods, U-Net and Segnet exhibit the highest levels of correctness and completeness. Our proposed methodology outperforms state-of-the-art methods in terms of quality metric and Dice. The quality metric, which represents the IoU of segmented images, demonstrates the superiority of our approach.

Figure 14 presents a visual comparison of state-of-the-art road detection methods, with the yellow box highlighting false positives for the road in the binary image. Basic Segnet achieves a low number of false positives while detecting occluded roads with continuity, as demonstrated in Segnet (b, c) of Figure 14. However, it struggles to identify certain road segments, as depicted in Segnet (d). In contrast, DR Segnet detects occluded roads without any discontinuity. The DRB, with its ability to aggregate more contextual information using varying convolution filter windows, improves the segmentation capability of occluded and junction roads. While U-net and ResU-net exhibit more discontinuities in the road due to occlusion, U-net fails to detect road continuity at junctions, as shown in U-net (a-d) of Figure 14. FCN detects the roads with very few false positives, as depicted in FCN (a, c, d) of Figure 14. It produces better visual results for occluded and junction roads compared to U-net and ResU-net, but suffers from memory consumption. The results of DR Segnet are superior to FCN, as it further enhances the continuity of road junctions.

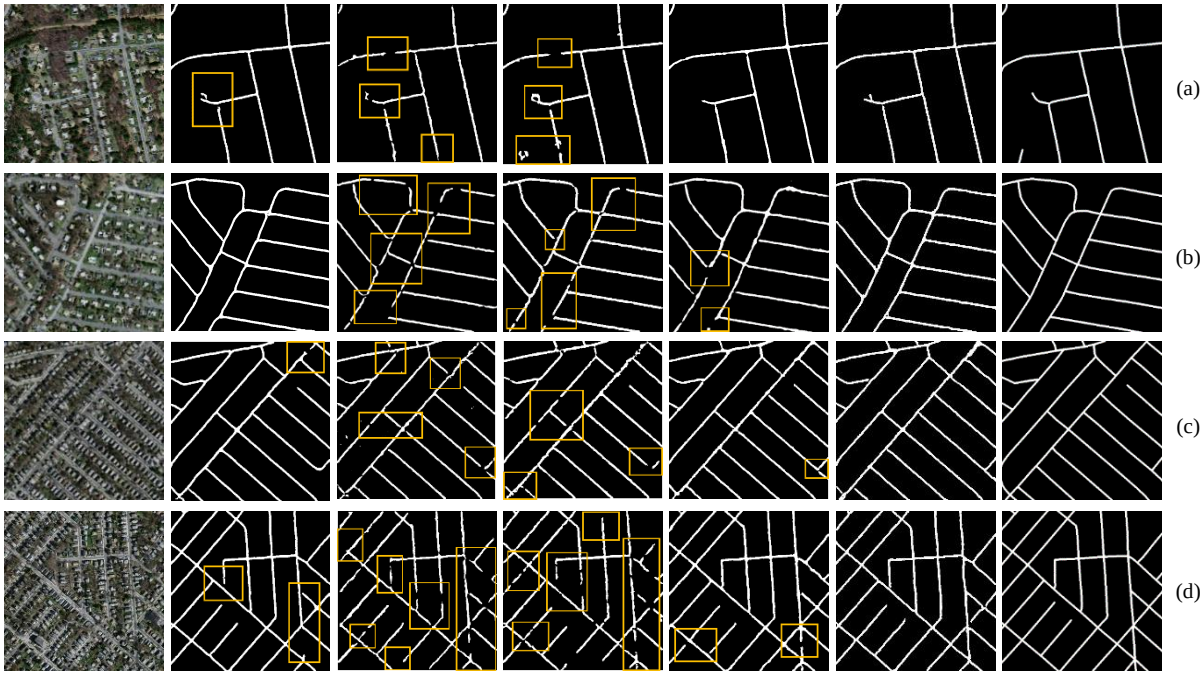

Figure 14: Simulation qualitative analysis on Massachusetts road dataset

Table 13: Comparison with published statistics on Massachusetts road dataset

Ref.	Methodology	Completeness	Dice	Correctness	Quality
[31]	Cascaded Attention Dense U	76.55	77.89	79.45	64.12
[32]	ResNet with Multi Task	72.54	78.13	84.65	64.11
[33]	Inter-Layer Self-Attention	56.66	66.19	79.58	-
[34]	Encoder-Decoder and DCRF	82.2	*84.6*	*87.1*	-
[35]	Dense U Net	70.41	74.07	78.25	**74.47**
[36]	Link Net	76.10	75.17	74.30	60.32
[37]	EU Net with Ensemble learning and Post Processing	81.31	80.45	80.71	68.56
[38]	Densely Connected Spatial Feature-Enhanced Pyramid	71.97	76.59	-	62.48
[39]	DA-Road Net	77.25	78.19	79.16	64.19
[40]	Deep FR TransNet	**83.72**	-	83.72	62.86
[41]	Richer U Net	-	-	-	62.77
Proposed Method		*82.94*	**85.30**	**87.87**	*74.41*

Table 13 shows the comparison of DenseResSegnet with other network. It clearly shows the superior performance of proposed network.

4.2.2 Ablation Study of DenseResSegnet

To verify the effectiveness of proposed method we have carried the ablation study of DR Segnet with fixed and variable number of convolution filters and dilated convolution with fixed

number of filters. Table 14 represents the mean values of performance metrics of 31 testing images. The DR Segnet outperforms in quality and Dice as compare to variable convolution filters in basic Segnet.

Table 14: Ablation study of proposed method with constant and variable filters and dilated convolution

Architecture	Completeness	Correctness	Quality	Dice	No. of epoch for convergence	Training Time
DR Segnet Dilated	*82.15*	85.01	71.73	83.47	72	2:09:42
DR Segnet with variable filter	82.11	*87.47*	*73.49*	*84.66*	*62*	2:21:04
Proposed	**82.94**	**87.87**	**74.41**	**85.30**	**63**	1:54:40

4.3. CONTRIBUTION 3- Semantic Segmentation building on Inria and Massachusetts dataset.

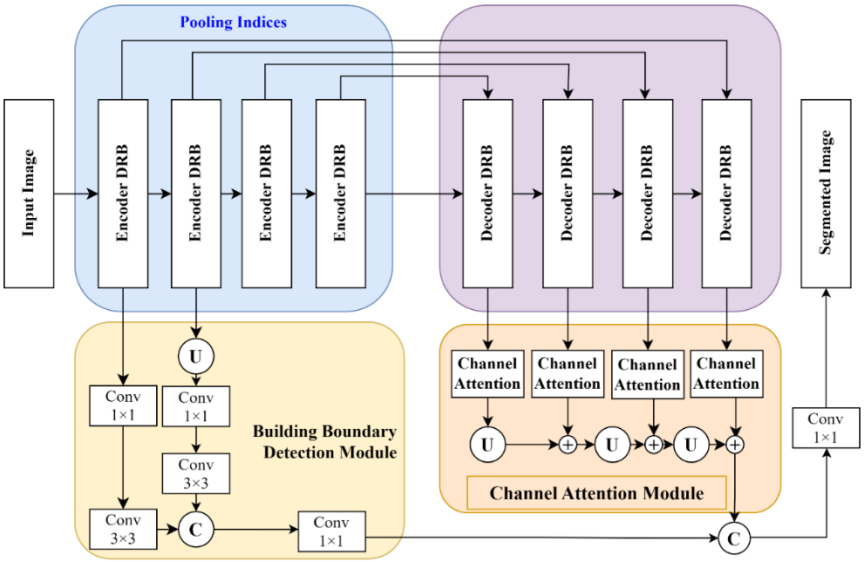

Figure 15: DenseResSegnet with building boundary detection and channel attention module

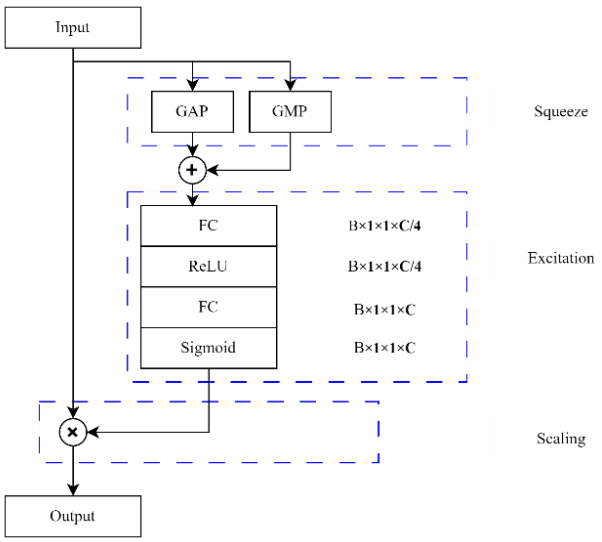

Figure 16: Squeeze and Excitation module as channel attention module [44].

Building boundaries plays crucial role in identifying the objects in the remote sensing images. As we know that the initial layers of CNN extract the low level features such edges in the images, so these features are extracted from initial layers from encoder side. This mechanism is called as building boundary detection module in Figure 15. These extracted features are concatenated with the channel attention module of Figure 16. The channel attention module is squeeze and excitation module, here in this we have used both global average and global max pooling for weighting each channel. The channel attention module discards the irrelevant channels.

4.3.1 Data set for simulation

Our proposed algorithm was validated using both the Inria [45] and Massachusetts building datasets. The Inria Building Dataset is a collection of high-resolution aerial images and their corresponding annotations of building footprints. We have made patches of 256 by 256 for training, validation and testing purpose. The images used for training are 1800 and testing 200. The other dataset for validating the proposed algorithm is Massachusetts building dataset which is a publicly available dataset of aerial images. It covers building footprints Worcester, Massachusetts, USA. The dataset includes orthorectified aerial images with a spatial resolution of 15 cm per pixel, captured in 2005 using a digital camera mounted on an aircraft. The dataset has also been used for benchmarking different algorithms and evaluating their performance.

Hyper parameter: For simulation we have used focal entropy loss with progressive learning rate with 100 epoch, batch size of 12, learning rate drop factor 0.1 with learning rate drop at 30 epochs, early stopping at with 7 epochs and initial learning rate of 0.001.

4.3.1 Results and Discussion on Building Detection.

Edge detection modules, is designed to detect edges and contours in images, which can be useful for accurately identifying the boundaries of the building. By fusing an edge detection module with DensResSegNet, the resulting model capture the boundaries of buildings also, hence improves segmentation results.

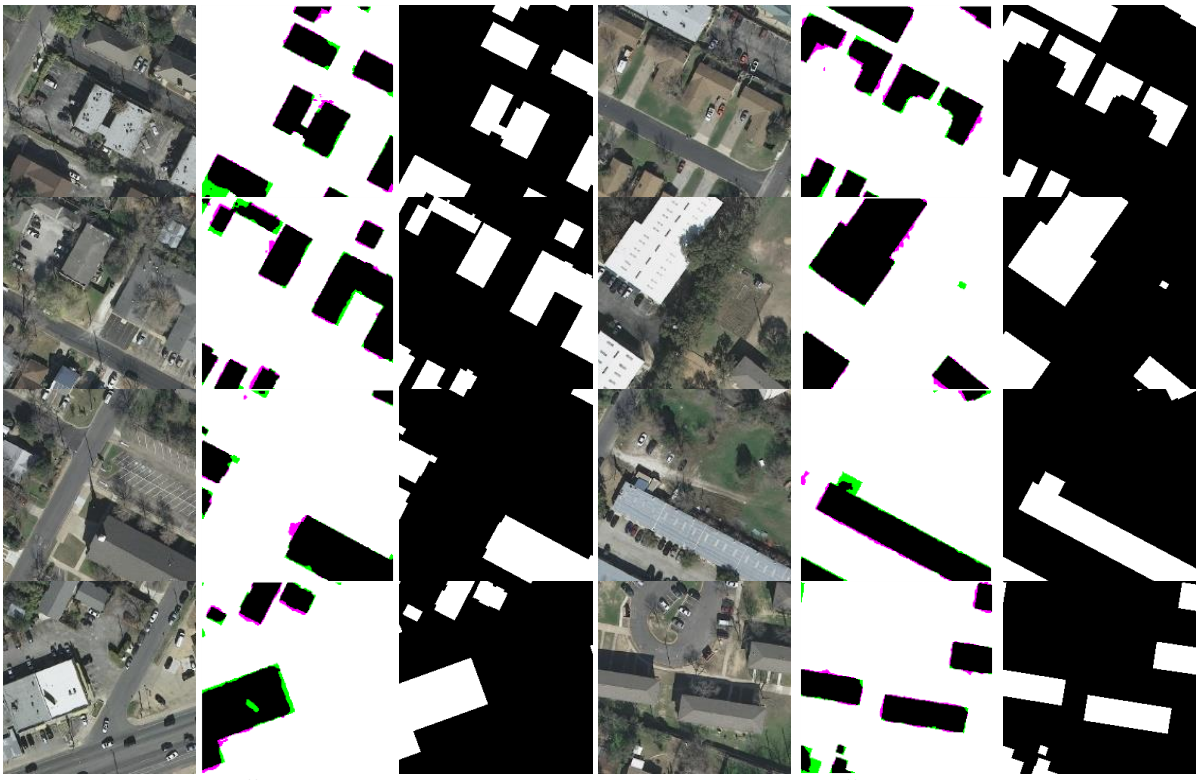

Image Difference Image Label Image Difference Image Label

Figure 17: Simulation results of proposed network Inria dataset.

Table 15: Performance metrics on Inria dataset

Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU
93.08	87.42	80.12	87.37

Table 16: Class wise performance metrics on Inria dataset

Class	Accuracy	IoU
building	78.26	68.37
Background	96.58	91.86

Purple and green color in segmented image of Figure 17 indicates the FP and FN pixels. By fusing building boundaries and channel attention improves the mean IoU of the building. The Massachusetts building data set results are shown in Table 15, Table 16 and Figure 18. The mIoU of testing images are 79.69.

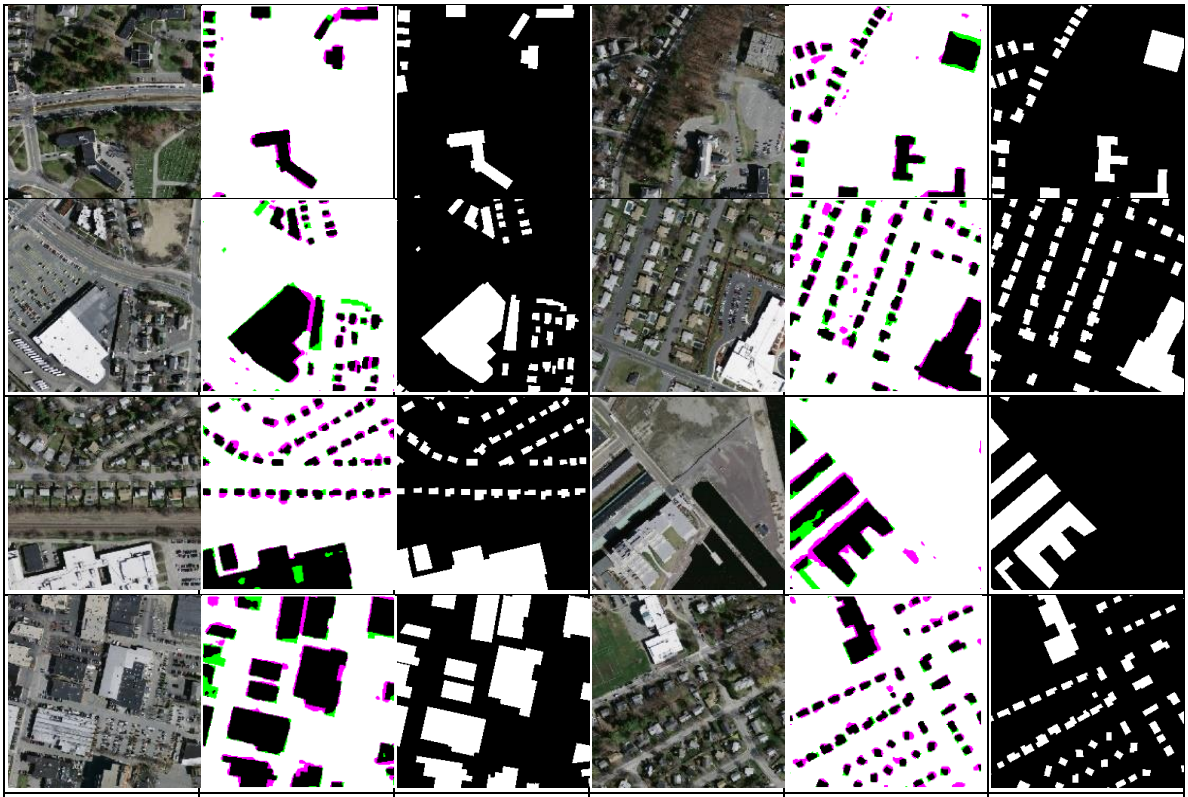

Figure 18: Simulation results on Massachusetts building dataset. Purple and green color indicates the FP and FN.

Table 17: Performance metrics on Massachusetts dataset

Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BF Score
93.81	89.45	79.69	89.05	84.63

Table 18: Class wise performance metrics on Massachusetts dataset

Class	Accuracy	IoU	Mean BF Score
building	83.27	66.44	81.58
Background	95.63	92.95	87.68

The performance of proposed network on Massachusetts building dataset are shown Table 17, Table 18 and Figure 18

4.4. CONTRIBUTION 4- Semantic Segmentation multiclass on Vaihingen dataset.

The factorization technique, first introduced in Inception-v3, has been utilized in ERF Net [47] to reduce the 3×3 convolutions in the original residual modules. Despite being more advantageous for larger filters, this decomposition has resulted in a 33% reduction in

parameters when applied to 3×3 convolutions, thus increasing the computational efficiency of the network.

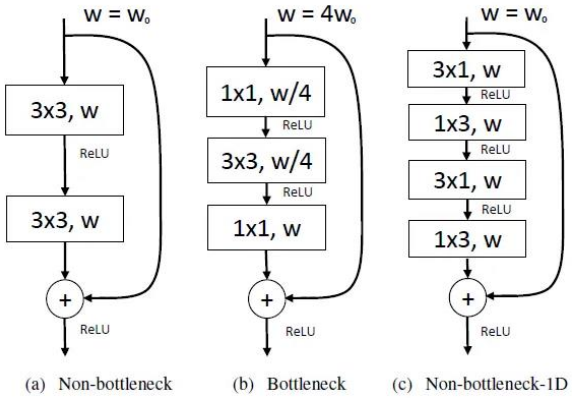

Figure 19: bottleneck layers with weight size

Table 19: Number of trainable parameters required for bottleneck layer

Residual block	Weights
Bottleneck	69K
Non-bottleneck	73K
Bottleneck-1D	57K
Non-bottleneck-1D	49K

Table 19 and Figure 19 summarizes the total dimensions of the weights on the convolutions of every residual block. The down sampling (Figure 20) we applied uses the combination of max pooling and strided convolution. The purpose of the bottleneck layer is to reduce the number of channels in the feature maps, which helps to reduce the computational complexity of the network. The smaller number of channels also makes it easier for the network to learn more efficient representations of the input data. The bottleneck layer is followed by a skip connection that allows the network to learn the residual between the input and output feature maps. The skip connection is a crucial component of the residual block, as it allows the network to learn residual features that are added back to the input feature maps. This helps to ensure that the network can learn deeper and more complex representations of the input data without sacrificing performance. Figure 21 shows the proposed bottleneck layer that uses cascaded of asymmetric convolution with residual and dense connection. The last dense shown by orange color uses the dilated convolution with dilation rate of 6. Figure 22 shows the bottleneck layer proposed in ERF net, we have used these layer with dilated convolution for parallel processing as shown in Figure 23.

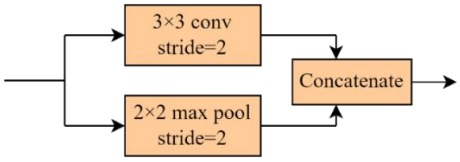

Figure 20: Down sampling scheme

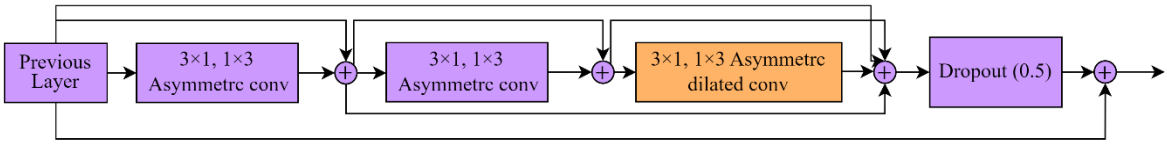

Figure 21: Proposed bottleneck layer

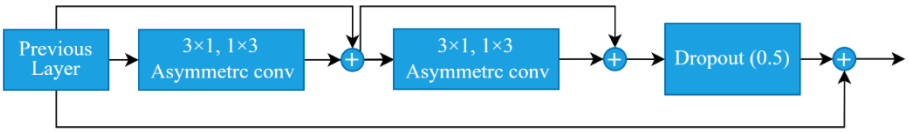

Figure 22: layers with dilated convolutions

Overall, the bottleneck layer in a key component that enables the network to achieve high accuracy while reducing computational complexity. Its use in the residual blocks of the network allows for efficient and effective feature extraction, making it a powerful tool for semantic segmentation tasks in computer vision.

4.4.1 Vaihingen and Deepglobe Dataset.

The International Society for Photogrammetry and Remote Sensing [46] provides semantic labelling benchmark Vaihingen dataset for semantic segmentation. It has 33 high-resolution true orthophoto (TOP) tiles, and GT labels. The size of these 33 tiles are of 2494×2064 with 9 cm ground sample distance. The 3 bands associated with each tile are near infrared, red, and green. The six classes in the data set are impervious surfaces, buildings, low vegetation, trees, cars and clutter. Out of six classes clutter has very limited pixels, hence we have skipped this and considered only 5 classes. The classes in the dataset are highly imbalanced, to address this we have used weighted cross entropy loss. The weighted cross entropy loss assigns a weight to each class based on its frequency in the dataset. The idea is to give more weight to the minority class and less weight to the majority class. This is done by multiplying the standard cross-entropy loss by the weight assigned to each class, which effectively balances the contribution of each class to the loss function. This way, the model can learn to give equal importance to all classes, regardless of their frequency in the dataset.

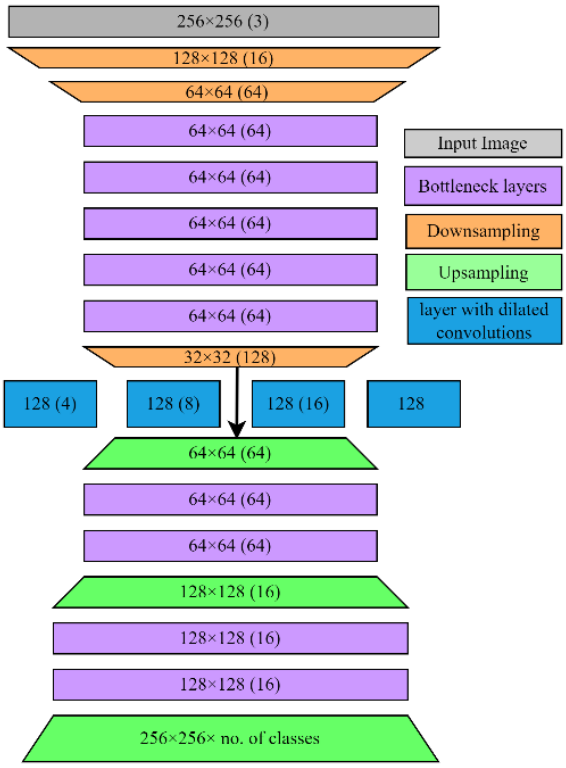

Figure 23: Proposed architecture for semantic segmentation

Hence by assigning higher weights to the underrepresented classes, the model is encouraged to pay more attention to these classes during training, and to minimize the loss for these classes more aggressively. This helps to achieve better overall performance, especially for imbalanced datasets.

Hyper parameter: For simulation we have used weighted cross entropy loss with progressive learning rate with 100 epoch, batch size of 24, learning rate drop factor 0.1 with learning rate drop at 30 epochs, early stopping at with 7 epochs and initial learning rate of 0.001.

Performance Metrics: The performance metrics used for evaluation of semantic segmentation are global accuracy, mean accuracy, mean IoU (mIoU), weighted IoU and mean BF score. Global accuracy refers to the percentage of pixels that are correctly classified, without taking into account the specific class to which they belong. The widely used metric to measure the accuracy of image segmentation is IoU (Figure 24), also known as the Jaccard similarity coefficient. This metric calculates the overlap between the predicted segmentation and the ground truth for each class. The mean IoU (mIoU) is the average IoU score of all classes in an individual image. On the other hand, the Mean IoU calculates the average IoU score across all classes for all images in the dataset. When dealing with images with unevenly sized classes, it is important to use the weighted average IoU of each class, based on the number of pixels in

that class. This approach helps to reduce the impact of errors in smaller classes on the overall quality score. The simulation results on Vaihingen dataset are shown in Table 20, Table 21 and Figure 25

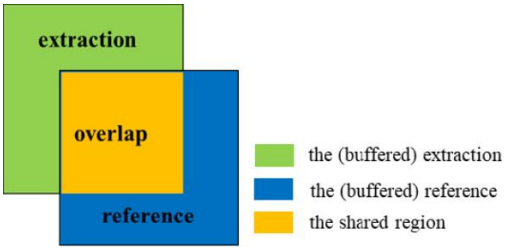

Figure 24: Intersection over Union [26]

Image	Proposed Net		ERF Net		Label
	Segmented output	Difference Image	Segmented output	Difference Image	

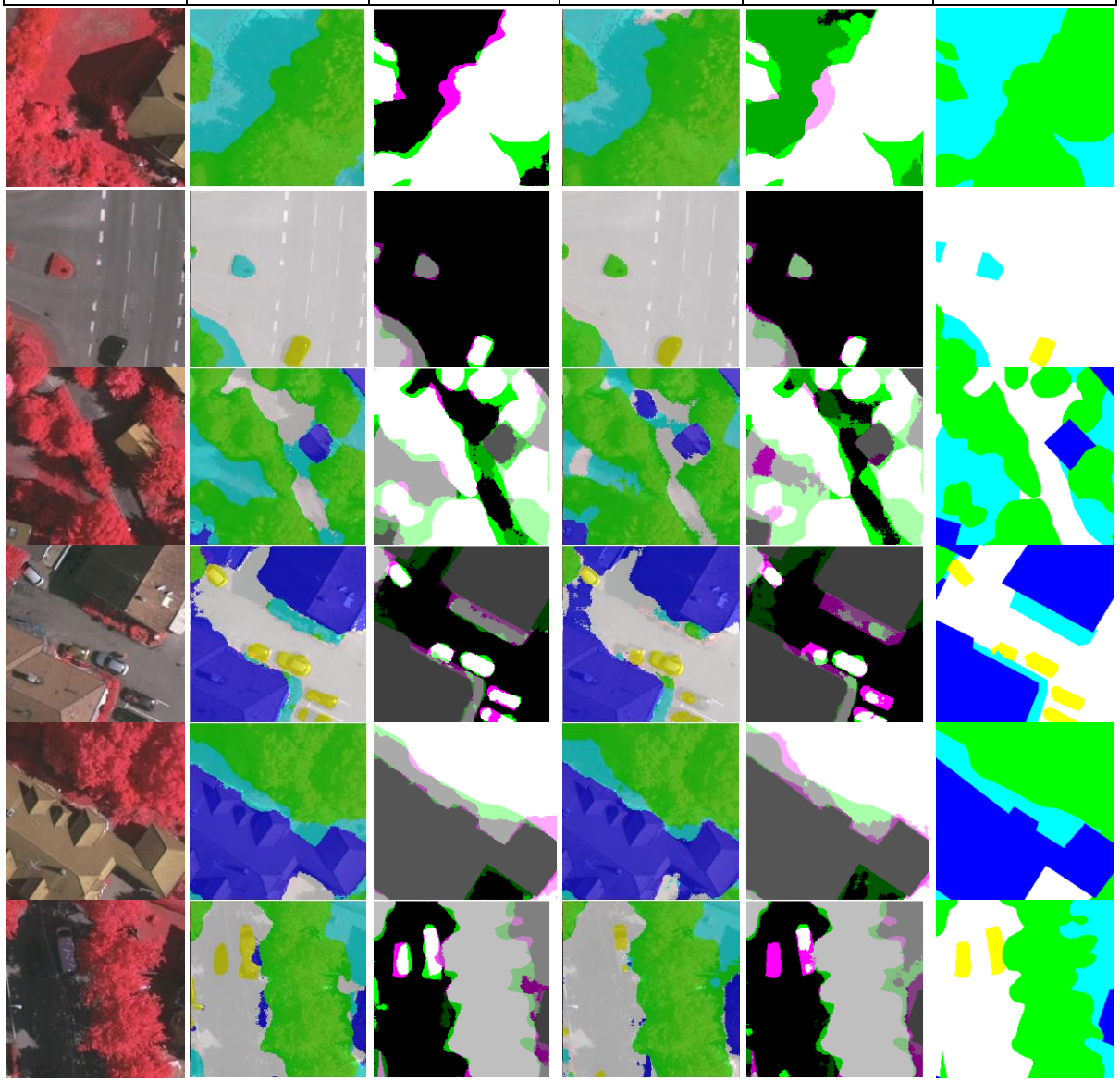

Figure 25: Visual comparison results of proposed network with ERF net

Table 20: Performance metrics on Massachusetts dataset

Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BF Score
87.66	87.39	73.24	78.39	50.11

Table 21: Class wise performance metrics on Massachusetts dataset

Class	Accuracy	IoU	Mean BF Score
road	88.09	83.77	52.44
building	94.53	86.43	61.88
low_vegetation	73.64	63.37	40.13
tree	92.83	78.67	48.20
car	87.84	53.97	45.41

4.4.2 Results and Discussion on Forest Detection.

Deep Globe Forest Detection dataset is a collection of high-resolution satellite images that have been annotated with labels for forest/non-forest regions. The dataset was created as part of the Deep Globe challenge, which aims to promote research and development in the field of satellite image analysis. The forest detection dataset contains a total of 803 training images and 137 validation images, each with a resolution of 2448×2448 pixels. This dataset is collected from the www.kaggle.com. The performance of the algorithm is shown in Table 22, Table 23 and Figure 26.

Table 22: Performance metrics on Deep Globe dataset

Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BF Score
88.66	87.15	78.44	79.56	51.93

Table 23: Class wise performance metrics on Deep Globe dataset

Class	Accuracy	IoU	Mean BF Score
Forest	93.95	83.50	59.27
Background	80.36	73.39	44.59

Image	Dense ERF net		DenseResSegETnet		Label
	Segmented output	Difference Image	Segmented output	Difference Image	

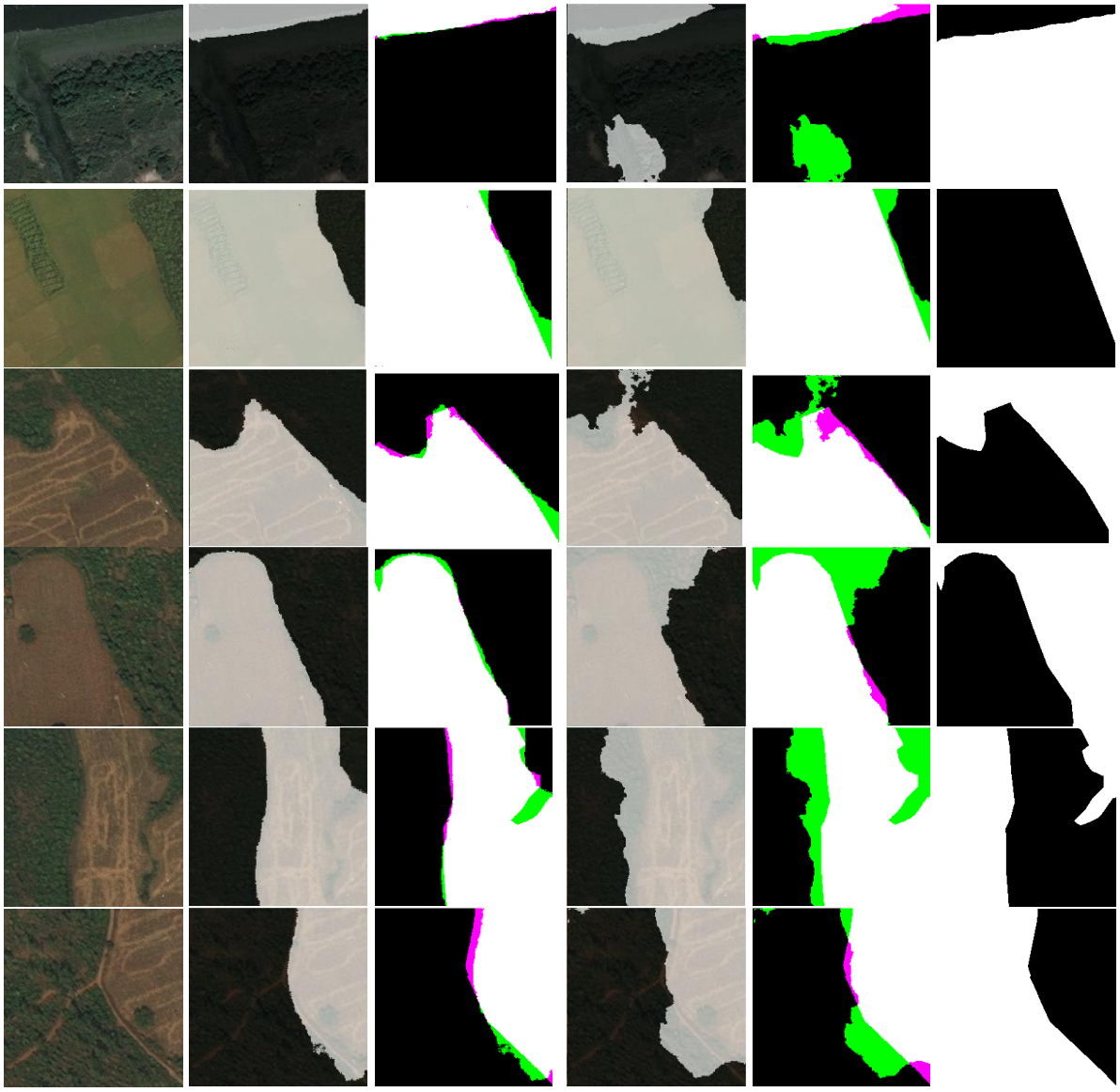

Figure 26: Simulation results on DeepGlobe forest data set.

CHAPTER 5

SUMMARY AND CONCLUSIONS

1. The zooming and shrinking capabilities of the features in model come from the fact that the receptive field of the neurons in the later layers becomes larger as the layer depth increases. By utilizing this zooming and shrinking capability of the features within the model, the model can effectively detect roads that vary in width.
2. Post-processing techniques can filter out small, non-road objects and join broken segments of roads in remote sensing images. As the deep learning algorithm continues to learn, it also attempts to learn more about background classes, leading to the shrinking of false positive pixels. However, these shrunk false positive pixels can be effectively filtered out during post-processing, ultimately leading to condensed training times.
3. By incorporating both residual and dense connections into the encoder and decoder, the deep learning algorithm can more effectively learn and identify patterns within the input data, ultimately improving its sensitivity and accuracy.
4. The computation training time of a deep learning algorithm is heavily dependent on the depth and number of filters utilized within the stages of both the encoder and decoder portions of the network. A greater depth and number of filters can lead to longer training times, while reducing these factors can result in shorter training times. Therefore, finding an optimal balance between depth and number of filters is crucial in order to achieve efficient training times for deep learning algorithms
5. The boundary edge attention guidance module is a powerful addition to object detection algorithms, providing additional contextual information that can improve the accuracy and reliability of detections.
6. The memory requirements of the GPU are dependent on the number of filters used in the convolution layer of the encoder and decoder in deep learning models. As the number of filters increases, so does the memory required to store their weights and perform calculations during training and inference. The equal number of filters in encoder and decoder section reduce this memory requirement and shortens the training time.
7. Asymmetrical convolution is a powerful technique for reducing memory consumption in deep learning models, allowing for larger batch sizes and smoother training curves. It can also improve the model's ability to capture different scales of features and lead to better performance in various tasks. A larger batch size means that the model can process more

examples in each iteration, which can lead to a more stable and accurate gradient estimation. This can result in faster convergence and better model performance.

8. The choice of loss function is a critical factor in training deep learning models, and the focal loss function has shown promising results in addressing the issue of class imbalance in RS datasets for binary class. By down-weighting the dominating class, the focal loss function can improve the performance of the model on the minority class and lead to better overall performance.
9. The performance of semantic segmentation models degrades with an increasing number of classes due to the complexity of the problem. Simple focal loss and existing architectures designed for binary classification problems are not well-suited for multi-class segmentation problems. To improve the performance of semantic segmentation models with an increasing number of classes, more advanced loss functions and modified architectures should be used.
10. The dilated convolution increases the field of view that helps in identifying the shape of the object. Insufficient field of view can result in suboptimal outcomes, especially when dealing with partial information about structures. This lack of information makes it difficult to distinguish between objects with similar textural features. In addition, predictions for pixels located at the edges of patches, can be unreliable due to the limited field of view.

CHAPTER 6
FUTURE WORK

In the future, the direction of semantic segmentation is likely to focus on improving the accuracy and efficiency of the segmentation process, particularly for two-class and multi-class segmentation. For two-class segmentation, the future direction will likely involve the development of more advanced techniques that can handle complex and overlapping object classes. The deep learning models have shown promising results in segmenting objects with high accuracy and robustness. However, there is still room for improvement, particularly in handling objects with similar appearance or occlusion. One possible direction for improving two-class segmentation is to incorporate additional information such as depth, motion, or texture cues into the segmentation process. These cues can provide complementary information to the visual features extracted from the image, which can help to disambiguate object boundaries and improve the accuracy of segmentation.

For multi-class segmentation, the future direction will likely focus on developing more efficient and scalable techniques that can handle a large number of object classes. Currently, many state-of-the-art techniques for multi-class segmentation rely on deep learning models such as encoder-decoder networks and attention-based models. However, these models can be computationally expensive and require large amounts of training data, which can limit their scalability and applicability in real-world scenarios. Therefore, future research may focus on developing more efficient and lightweight models that can handle a large number of object classes while maintaining high accuracy.

Additionally, research may also focus on developing techniques that can handle imbalanced datasets, where some object classes may have significantly fewer examples than others. This is particularly important in real-world scenarios where some object classes may be rare or difficult to obtain in large quantities. In summary, the future direction of semantic segmentation for both two-class and multi-class segmentation is likely to focus on improving the accuracy, efficiency, and scalability of the segmentation process, as well as developing techniques that can handle complex and imbalanced datasets.

PUBLICATION

1. Akhtar, N., Choubey, N. S., & Ragavendran, U. (2019). Investigation of Non-natural Information from Remote Sensing Images: A Case Study Approach. Computational Intelligence and Sustainable Systems: Intelligence and Sustainable Computing, 165-199. (**Scopus**)
2. Akhtar, N., & Ragavendran, U. (2020). Interpretation of intelligence in CNN-pooling processes: a methodological survey. Neural computing and applications, 32(3), 879-898. (**Scopus, SCIE, IF: 4.77, Q1**).
3. Nadeem Akhtar, Manish Mandloi, (2023), DenseResSegnet: A Dense Residual Segnet for Road Detection Using Remote Sensing Images, IEEE conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS) 2023 Co-Sponsored by IEEE Geoscience and Remote Sensing Society (IGRSS) held during 27-29 Jan 2023. (**Unpublished Scopus**)

UNDER REVIEW

1. Road detection from aerial images with multi scale feature extraction and dilated inception, submitted to International Journal of Applied Remote Sensing.

REFERENCES

1. Shunta Saito; Ryota Arai; Yoshimitsu Aoki, "Seamline Determination Based on Semantic Segmentation for Aerial Image Mosaicking," *in IEEE Access*, vol. 3, pp. 2847-2856, 2015, doi: 10.1109/ACCESS.2015.2508921.
2. B. Yu, L. Yang and F. Chen, "Semantic Segmentation for High Spatial Resolution Remote Sensing Images Based on Convolution Neural Network and Pyramid Pooling Module," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 9, pp. 3252-3261, Sept. 2018, doi: 10.1109/JSTARS.2018.2860989.
3. Michele Volpi, and Devis Tuia. "Deep multi-task learning for a geographically-regularized semantic segmentation of aerial images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 144, pp. 48–60, 2018, https://doi.org/10.1016/j.isprsjprs.2018.06.007
4. Duc My Vo, and Sang-Woong Lee, "Semantic image segmentation using fully convolutional neural networks with multi-scale images and multi-scale dilated convolutions, "*Multimedia Tools and Applications*", vol. 77, no. 14, pp. 18689–18707, 2018, https://doi.org/10.1007/s11042-018-5653-x
5. Cheng Peng, Yangyang Li, Licheng Jiao, Yanqiao Chen, Ronghua Shang, "Densely Based Multi-Scale and Multi-Modal Fully Convolutional Networks for High-Resolution Remote-Sensing Image Semantic Segmentation," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 8, pp. 2612-2626, Aug. 2019, doi: 10.1109/JSTARS.2019.2906387.
6. Haifeng Luo, Chongcheng Chen, Lina Fang, Xi Zhu, Lijing Lu, "High-Resolution Aerial Images Semantic Segmentation Using Deep Fully Convolutional Network with Channel Attention Mechanism," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 9, pp. 3492-3507, Sept. 2019, doi: 10.1109/JSTARS.2019.2930724.
7. Azimi, Seyed Majid, Peter Fischer, Marco Körner, and Peter Reinartz. "Aerial LaneNet: Lane-marking semantic segmentation in aerial imagery using wavelet-enhanced cost-sensitive symmetric fully convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 5, pp. 2920–2938, 2019, doi: 10.1109/TGRS.2018.2878510.
8. Zhiling Guo, Guangming Wu, Xiaoya Song, Wei Yuan, Qi Chen, Haoran Zhang, Xiaodan Shi, Mingzhou Xu, Yongwei Xu, Ryosuke Shibasaki, Xiaowei Shao. "Super-Resolution Integrated Building Semantic Segmentation for Multi-Source Remote Sensing Imagery," in *IEEE Access*, vol. 7, pp. 99381-99397, 2019, doi: 10.1109/ACCESS.2019.2928646.
9. Ming Wu, Chuang Zhang, Jiaming Liu, Lichen Zhou, Xiaoqi Li. "Towards Accurate High Resolution Satellite Image Semantic Segmentation," in *IEEE Access*, vol. 7, pp. 55609-55619, 2019, doi: 10.1109/ACCESS.2019.2913442.

10. Nicolas Audebert, Alexandre Boulch, Bertrand Le Saux, and Sébastien Lefèvre. “Distance transform regression for spatially-aware deep semantic segmentation”, *Computer Vision and Image Understanding*, vol. 189, Article ID 102809, 2019, https://doi.org/10.1016/j.cviu.2019.102809.
11. L. Ding, T. Tang, and L. Bruzzone, “Improving semantic segmentation of aerial images using patch-based attention,” 2019, https://arxiv.org/abs/1911.08877
12. Teerapong Panboonyuen, Kulsawasd Jitkajornwanich, Siam Lawawirojwong, Panu Srestasathiern, Peerapon Vateekul, “Semantic segmentation on remotely sensed images using an enhanced global convolutional network with Channel attention and domain specific transfer learning,” *Remote Sensing*, vol. 11, no. Issue. 1, p. 83, 2019, https://doi.org/10.3390/rs11010083.
13. Wenjie Liu, Yongjun Zhang, Haisheng Fan, Yongjie Zou, Zhongwei Cui, "A New Multi-Channel Deep Convolutional Neural Network for Semantic Segmentation of Remote Sensing Image," in *IEEE Access*, vol. 8, pp. 131814-131825, 2020, doi: 10.1109/ACCESS.2020.3009976.
14. Fang Wang, Shihao Piao, Jindong Xie, "CSE-HRNet: A Context and Semantic Enhanced High-Resolution Network for Semantic Segmentation of Aerial Imagery," in *IEEE Access*, vol. 8, pp. 182475-182489, 2020, doi: 10.1109/ACCESS.2020.3028445.
15. Ruigang Niu, Xian Sun, Yu Tian, Wenhui Diao, Kaiqiang Chen, Kun Fu, Hybrid Multiple Attention Network for Semantic Segmentation in Aerial Images," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-18, 2022, Art no. 5603018, doi: 10.1109/TGRS.2021.3065112.
16. Wuttichai Boonpook, Yumin Tan, and Bo Xu, “Deep learning-based multi-feature semantic segmentation in building extraction from images of UAV photogrammetry”, *International Journal of Remote Sensing*, vol. 42, no. Issue 1, pp. 1–19, 2021, https://doi.org/10.1080/01431161.2020.1788742
17. Jianfeng Huang, Xinchang Zhang, Ying Sun, Qinchuan Xin, "Attention-Guided Label Refinement Network for Semantic Segmentation of Very High-Resolution Aerial Orthoimages," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4490-4503, 2021, doi: 10.1109/JSTARS.2021.3073935.
18. Muhammad Alam, Jian-Feng Wang, Cong Guangpei, LV Yunrong & Yuanfang Chen, “Convolutional neural network for the semantic segmentation of remote sensing images”, *Mobile Networks and Applications”*, vol. 26, no. 1, pp. 200–215, 2021, https://doi.org/10.1007/s11036-020-01703-3
19. Athanasios Anagnostis, Aristotelis C. Tagarakis, Dimitrios Kateris, Vasileios Moysiadis, Claus Grøn Sørensen, Simon Pearson “Orchard mapping with deep learning semantic segmentation,” *Sensors*, vol. 21, no. 11, 2021, https://doi.org/10.3390/s21113813.

20. Fang Wang, Xiaoyan Luo, Qixiong Wang, and Lu Li, “Aerial-BiSeNet: A real-time semantic segmentation network for high resolution aerial imagery”, *Chinese Journal of Aeronautics*, vol. 34, no. 9, pp. 47–59, 2021, https://doi.org/10.1016/j.cja.2020.10.032.
21. Zixian Zheng, Xueliang Zhang, Pengfeng Xiao, Zhenshi Li, "Integrating Gate and Attention Modules for High-Resolution Image Semantic Segmentation," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4530-4546, 2021, doi: 10.1109/JSTARS.2021.3071353.
22. Song Ouyang, Yansheng Li, “Combining deep semantic segmentation network and graph convolutional neural network for semantic segmentation of remote sensing imagery,” *Remote Sensing*, vol. 13, no. 1, p. 119, 2021, https://doi.org/10.3390/rs13010119
23. Mnih, V. Machine learning for aerial image labeling. Ph. D. dissertation, University of Toronto (Canada), 2013.
24. Xiaoyan Lu, Yanfei Zhong, Zhuo Zheng, Liangpei Zhang, “GAMSNet: Globally aware road detection network with multi-scale residual learning”, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 175, pp. 340-352, 2021, https://doi.org/10.1016/j.isprsjprs.2021.03.008
25. Li Chen, Qing Zhu, Xiao Xie, Han Hu, Haowei Zeng, “Road extraction from VHR remote-sensing imagery via object segmentation constrained by Gabor features”, *ISPRS international journal of geo-information*, vol. 7, no. 9, 362, 2018. https://doi.org/10.3390/ijgi7090362.
26. Xiaoyan Lu, Yanfei Zhong, Zhuo Zheng, Yanfei Liu, Ji Zhao, Ailong Ma, Jie Yang, "Multi-Scale and Multi-Task Deep Learning Framework for Automatic Road Extraction," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9362-9377, Nov. 2019, doi: 10.1109/TGRS.2019.2926397.
27. Vijay Badrinarayanan, Alex Kendall and Roberto Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481-2495, 1 Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.
28. Olaf Ronneberger, Philipp Fischer & Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science (), vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28
29. Zhengxin Zhang, Qingjie Liu, Yunhong Wang, "Road Extraction by Deep Residual U-Net," in *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749-753, May 2018, doi: 10.1109/LGRS.2018.2802944.

30. Evan Shelhamer, Jonathan Long, Trevor Darrell, T. “Fully convolutional networks for semantic segmentation”, *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 4, pp. 640-651, 2016. https://doi.org/10.1109/TPAMI.2016.2572683
31. Jing Li, Yong Liu, Yindan Zhang, Yang Zhang, “Cascaded Attention DenseUNet (CADUNet) for Road Extraction from Very-High-Resolution Images”, *ISPRS International Journal of Geo-Information*, vol. 10, no. 5, pp. 329, 2021, https://doi.org/10.3390/ijgi10050329
32. Ruonan Chen, Xiang Li, Yuan Hu, Congcong Wen and Ling Peng, "Road Extraction from Remote Sensing Images in Wildland–Urban Interface Areas," in *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1-5, 2022, Art no. 3000705, doi: 10.1109/LGRS.2020.3028468.
33. Jiahai Tan, Ming Gao, Kai Yang, and Tao Duan. "Remote sensing road extraction by road segmentation network." *Applied Sciences*, vol. 11, no. 11, 2021, pp. 5050, https://doi.org/10.3390/app11115050
34. Shuyang Wang, Xiaodong Mu, Dongfang Yang, Hao He, and Peng Zhao. "Road extraction from remote sensing images using the inner convolution integrated encoder-decoder network and directional conditional random fields." *Remote Sensing*, vol. 13, no. 3, pp. 465, 2021, https://doi.org/10.3390/rs13030465
35. Jiang Xin, Xinchang Zhang, Zhiqiang Zhang, Wu Fang, “Road extraction of high-resolution remote sensing images derived from DenseUNet”, *Remote Sensing*, vol. 11, no. 21, pp. 2499, https://doi.org/10.3390/rs11212499
36. Xungen Li, Zhan Zhang, Shuaishuai Lv, Mian Pan, Qi Ma, Haibin Yu. "Road Extraction From High Spatial Resolution Remote Sensing Image Based on Multi-Task Key Point Constraints," in *IEEE Access*, vol. 9, pp. 95896-95910, 2021, doi: 10.1109/ACCESS.2021.3094536.
37. Junjie Li, Yizhuo Meng, Donyu Dorjee, Xiaobing Wei, Zhiyuan Zhang, Wen Zhang. "Automatic Road Extraction from Remote Sensing Imagery Using Ensemble Learning and Postprocessing," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 10535-10547, 2021, doi: 10.1109/JSTARS.2021.3094673.
38. Qiangqiang Wu, Feng Luo, Penghai Wu, Biao Wang, Hui Yang, Yanlan Wu, "Automatic Road Extraction from High-Resolution Remote Sensing Images Using a Method Based on Densely Connected Spatial Feature-Enhanced Pyramid," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3-17, 2021, doi: 10.1109/JSTARS.2020.3042816.
39. Jie Wan; Zhong Xie, Yongyang Xu, Siqiong Chen, Qinjun Qiu, "DA-RoadNet: A Dual-Attention Network for Road Extraction from High Resolution Satellite Imagery," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 6302-6315, 2021, doi: 10.1109/JSTARS.2021.3083055.
40. Zhijin Ge, Yanling Zhao, Jin Wang, Duo Wang, Qi Si, "Deep Feature-Review Transmit Network of Contour-Enhanced Road Extraction from Remote Sensing Images," in *IEEE Geoscience and*

Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 3001805, doi: 10.1109/LGRS.2021.3061764.

41. Yifan Zao; Zhenwei Shi, "Richer U-Net: Learning More Details for Road Detection in Remote Sensing Images," in *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1-5, 2022, Art no. 3003105, doi: 10.1109/LGRS.2021.3081774.
42. Nadeem A., Manish M., DenseResSegnet: A Dense Residual Segnet for Road Detection Using Remote Sensing Images, IEEE conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS) 2023 Co-Sponsored by IEEE Geoscience and Remote Sensing Society (IGRSS) held during 27-29 Jan 2023.
43. Hongzhang Xu, Hongjie He, Ying Zhang, Lingfei Ma, Jonathan Li, "A comparative study of loss functions for road segmentation in remotely sensed road datasets*", International Journal of Applied Earth Observation and Geoinformation*, 116, 103159, 2023, doi: 10.1016/j.jag.2022.103159
44. Jie Hu; Li Shen; Gang Sun, "Squeeze-and-Excitation Networks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 7132-7141, doi: 10.1109/CVPR.2018.00745.
45. Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, Pierre Alliez, "Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark," 2017 *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Fort Worth, TX, USA, 2017, pp. 3226-3229, doi: 10.1109/IGARSS.2017.8127684.
46. https://www.isprs.org/publications/archives.aspx
47. Eduardo Romera, José M. Álvarez, Luis M. Bergasa and Roberto Arroyo, "ERFNet: Efficient Residual Factorized ConvNet for Real-Time Semantic Segmentation," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 263-272, Jan. 2018, doi: 10.1109/TITS.2017.2750080.